Course Overview

Surveys of Bioinformatics
The SYBB Survey Series is composed of the following course sequence: (A) Technologies in Bioinformatics, (B) Data Integration in Bioinformatics, (C) Translational Bioinformatics, and (D) Programming for Bioinformatics. Each standalone section of this course series introduces students to an aspect of a bioinformatics project - from data collection (SYBB 311/411A), to data integration (SYBB 311/411B), to research applications (SYBB 311/411C), with a fourth module (SYBB 311/411D) introducing basic bioinformatics programming skills. Each semester SYBB 311/411A, SYBB 311/411B, and SYBB 311/411D offered as blocks meeting back to back. SYBB 311/411C is a longitudinal class throughout the semester. Each class is graded separately.

Course Description
SYBB 311D/411D is a 1 credit, 5-week long course that will introduce students to bioinformatics software and programming in the R language; this course is designed for those with little or no prior programming experience. However, advanced programmers can still learn bioinformatics pipelines and software packages to conduct research. Students will gain hands-on experience working with R packages and functions designed for bioinformatics applications.

Programming for Bioinformatics short course focuses on a platform, in this case R-project (rproject.org), and introduces students to basic programming in R, what packages are available for their use, and teaches an introductory hands-on experience working with R by walking through the students in analyzing a large –omics dataset. At the end of the class, the students are assessed with a small-scale project, where they analyze a publicly available dataset and produce a short report.

Course Objectives

<table>
<thead>
<tr>
<th>Big Idea</th>
<th>Enduring Understandings</th>
<th>Learning Outcomes</th>
</tr>
</thead>
</table>
| Data Analysis in Bioinformatics | Understand Bioinformatics tools and models to build pipelines to interpret and analyze large datasets | • Can assess tools and packages provided for bioinformatics analysis  
• Can summarize and represent datasets  
• Complete basic -omics data analysis (e.g. differential gene expression) |
| | Develop software solutions to bioinformatics problems | • Use open source bioinformatics software/platforms (R, Cytoscape)  
• Understand fundamental programming concepts  
• Manipulate data-files to prepare for bioinformatics analysis |
**Course Prerequisites**

**Undergraduate**  
**Prerequisites:** BIOL 214 AND 215; OR 250  
**Co-requisites:** SYBB311A, SYBB311B and SYBB311D

**Graduate**  
**Prerequisites:** Graduate standing OR Prerequisites not met permission  
Graduate students have the option of enrolling in all four courses or choosing the individual modules most relevant to their background and goals with the exception of SYBB411D, which must be taken with SYBB411A. Based on experience in the field, instructor can waive this co-requisite.

**Required Texts and Materials**

**Online Textbook(s)**  
This is an e-book that is free to access. Please make sure you can access to this resources.

**Course Website:**  
The course website is available on Blackboard. The URL is: http://blackboard.case.edu. Students will need their CWRU network ID and password to access the site. Important course information and announcements will be posted on Blackboard including the course syllabus, assignments, and grades. The students are responsible to refer to this site regularly. Additional information on the SYBB 311/411 Series can be found on Instructors web page http://gurkan.case.edu/teaching.html.

**Course format**  
This is an active-learning class. The focus is on understanding and critically evaluating bioinformatics analysis approaches and building basic programming skills. Readings, tutorials and in class exercises will be assigned on a weekly basis. Students are expected to complete these assignments before coming to class. There will be homework assignments and a class project to assess students’ understanding of the material. The assignments will include hands-on work and basic problem solving exercises.

In class exercises can be assigned before class. These will be shared on blackboard. Students should follow the directions, and provide written evidence of completion via Blackboard when asked.

**Course Policies**

**General**

1. Students are asked to bring personal computers. If not available, a lab computer will be assigned to students.

2. Office hours are open to all. However, emailing the instructor/T.A. ahead of time is suggested.

3. Course materials, including assignments will be posted on Blackboard. Students are responsible to gain access to Blackboard, and complete these assignments on time.

**Assignments**

1. Students are expected to work independently. Offering and accepting solutions from others is an act of plagiarism, which is a serious offense and all involved parties will be penalized according to the Academic Honesty Policy. Discussion amongst students is encouraged, but when in doubt, direct your questions to the professor, or lab assistant.

2. No late assignments will be accepted under any circumstances.

3. All students are expected to complete the class project.
Attendance and Absences

1. Attendance in full is expected. Attendance will not be taken.
2. Students are responsible for all missed work, regardless of the reason for absence. It is also the absentee’s responsibility to get all missing notes or materials.

Participation

1. Participation in class means evidence for involvement in meaningful discussion.
2. Students have to show a presence for full participation credit.
3. A physical presence is not sufficient to earn credit.
4. Several research concepts, papers and R packages will be covered in the class. In each class session, the students are expected to participate in class discussions. The students should be ready to deliberate about the reading assignments (see Course Policies).

Computer Use

1. Computer use is strictly limited for class materials only.
2. Any use of computers, tablets, phones that are not related to the course material (e.g. social networking, texting etc.) is not allowed. This will result in a zero grade from class participation (0/10).

Course Outline:
The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments.

<table>
<thead>
<tr>
<th>Week/Lecture</th>
<th>Topic</th>
<th>Content (Topics/Readings/Assignments)</th>
</tr>
</thead>
</table>
| Lecture 1    | Introduction to R and Data Structures | • Introduction to R  
• Introduction to data structures and variables.  
• Assignment #1 will be handed. |
| Lecture 2    | Data Structures and Conditional Constructs and Functions | • Solutions for Assignment #1 will be discussed.  
• Data structures  
• If-else conditional constructs  
• Functions  
• Assignment #2 will be handed. |
| Lecture 3    | Microarray Data Analysis | • Analysis and comprehension of high-throughput data with Bioconductor.  
• Reading Assignment:  
• Class projects are handed out. |
| Lecture 4    | Iterative Loops and NGS Data Analysis | • Solutions for Assignment #2 will be discussed.  
• Basic concepts in programming (Combining Boolean and Comparison Operators, Iterative Constructs (Loops)).  
• Next Generation Sequencing Data Analysis (RNA-seq)  
• Assignment #2 will be handed.  
<table>
<thead>
<tr>
<th>Week/Lecture</th>
<th>Topic</th>
<th>Content (Topics/Readings/Assignments)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture 5</td>
<td>Biological Sequence Analysis</td>
<td>• Computational analysis of biological sequences.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Bioconductor Package Resources</td>
</tr>
<tr>
<td>Lecture 6</td>
<td>Data Management</td>
<td>• Solutions for Assignment #2 will be discussed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Basic concepts in programming (Data management, input/output).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Computational analysis of biological sequences (contd.).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Assignment #3 will be handed.</td>
</tr>
<tr>
<td>Lecture 7</td>
<td>R Lab</td>
<td>• R, R-studio and Bioconductor Package installation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Revise Microarray Data Analysis (lumi and affy packages)</td>
</tr>
<tr>
<td>Lecture 8</td>
<td>R-Lab</td>
<td>• Class Project Overview</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Data manipulation for data analysis</td>
</tr>
<tr>
<td>Lecture 9</td>
<td>Biological Networks</td>
<td>• Interaction networks, and basic network analysis.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Data manipulation in R for building Networks.</td>
</tr>
<tr>
<td>Lecture 10</td>
<td>Visualization in R</td>
<td>• Solutions for Assignment #3 will be discussed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Plotting graphs and tables.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Class projects are due before the first Saturday after the last lecture.</td>
</tr>
</tbody>
</table>

**Homework**

(All students) The instructor will hand out 4-5 practical assignments. In each assignment, one or more problems will be provided, and the students will be asked to write code or analyze data to answer specific questions. The assignments will be posted on Blackboard, and announced in class. Students will submit their results via Blackboard. No paper submission will be accepted unless noted.

**Final Group Project**

The goal of the class project is to carry out reproducible bioinformatics analysis of a large —omics data set. The students will either download datasets suggested or utilize their own data (approved by the instructor) to complete the project.

The projects will be conducted in groups. Students taking SYBB 411D will carry out a more extensive analysis. The instructor will handout a list of datasets that can be accessed from public resources. The student will download, process, and analyze this dataset as part of the project.

- Each group of student will be given a separate dataset. Students that wish to utilize their own dataset should confirm with the Instructor by end of second lecture.
- SYBB 311D students will only generate a report of significant genes identified from the analysis. However, the SYBB 411D students are required to pursue a network analysis following the initial analysis.
- **Project report:** Each group will submit a project report. The report will outline the motivation and clearly state the steps taken to finish the analysis. Significant results should be reported at the end. The report is due before the next Saturday after the last lecture. Late assignments **will not be accepted**.
- **Grading:** The class project will be evaluated by the report submitted. The reports should be submitted in .pdf or .doc format to the instructor via blackboard (not email) with relevant additional files (no data files, no compressed folders.). The reports should highlight the steps taken. The following table describes the rubric for grading:
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Undergraduate Grading Scale</th>
<th>Graduate Grading Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of the data analysis</td>
<td><strong>All:</strong> Provide evidence of your analysis. Use minimal number of manual steps. Use appropriate packages and methods. Draw meaningful results from the datasets. Summarize your results in tables and plots.</td>
<td>33%</td>
<td>25%</td>
</tr>
<tr>
<td>Organization, reporting, and conclusions</td>
<td><strong>All:</strong> Provide a short report (not more than ~2000 words excluding your tables, references and 16% figure legends). The report should cover the logic followed in analysis, and the steps taken. Present any findings with tables and plots. Please use R to generate your plots. Plots, tables, and summaries generated with spreadsheet software will not be accepted. Discuss your final results briefly.</td>
<td>33%</td>
<td>25%</td>
</tr>
<tr>
<td>Graduate students:</td>
<td><strong>Describe the biomedical relevance of the project. Include plots that extend the analysis to patient oriented research (e.g. survival analysis).</strong></td>
<td>-</td>
<td>25%</td>
</tr>
<tr>
<td>Generate and report reproducible research</td>
<td><strong>All:</strong> Provide steps for data download, manipulation, results, plots, tables and relevant textual statements relating to these. Provide your source code, scripts and any additional data files not publicly available for replication of the study. Follow Reading Assignment #1 as a guideline and for more details.</td>
<td>33%</td>
<td>25%</td>
</tr>
</tbody>
</table>

**Grading**

Grading will be based primarily of the homework assignments, of which there will be 2 per week. An additional piece of your grade will come from participation in Blackboard and in-class discussions. Though attendance is not enforced, class participation will be evaluated as follows:

<table>
<thead>
<tr>
<th>Assignments</th>
<th>Undergraduate %</th>
<th>Graduate %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignments</td>
<td>40 %</td>
<td>30 %</td>
</tr>
<tr>
<td>Class Project</td>
<td>50 %</td>
<td>65 %</td>
</tr>
<tr>
<td>Participation</td>
<td>10 %</td>
<td>5 %</td>
</tr>
</tbody>
</table>
Final letter grades will correspond to the percentage of points you earn out of the possible 100 point. The tentative scale is $\geq 90\%=A$, $80-89.4\%=B$, etc. However, these ranges may be adjusted at the discretion of the instructor.

**Make-up Policy:**
Schedule of classes are listed on this document. Requests for make-up will only be considered in the event of an emergency that can be verified (e.g., a doctor’s note must be provided & possible verification from Undergraduate Studies) or a pre-scheduled academic conflict (presenting at a seminar series, presenting at a conference (Students have to notify instructor at least 2 weeks prior to this conflict).

If a student misses final paper submission due to an illness or other emergency, students are responsible to contact right away to report their emergency (48 hours). If the student fails to do this, they will not be granted a make-up and they will earn a zero for the paper.

**Disabilities**
If you have a disability and anticipate needing accommodations for this course, we are willing to work with you and Disability Resources (http://students.case.edu/education/disability) to help provide the accommodations you need.

**Academic Integrity**
No form of academic dishonesty including cheating, plagiarism, misrepresentation, or obstruction will be tolerated in this class. Plagiarism in any form will result in a failing grade. Please see: http://studentaffairs.case.edu/office/judicial

Students are expected to abide by the academic integrity policy of the university, which can be found at: http://studentaffairs.case.edu/handbook/policy/integrity.html. It is the student’s responsibility to read and familiarize him/herself with the academic integrity policy.

Evidence of academic misconduct will be reported to the Dean and could result in a judicial action. If a student is caught cheating, they may fail the assignment or the entire course depending on the severity of the offense. Examples of academic integrity violations include:

1. Possession of a cell phone, calculator or tablet during an exam.
2. Copying from a student’s exam.
3. Lying about an illness or other emergency to get out of taking an exam.
4. Participating in class with another student’s clicker.

**Data for Research Disclosure**
Any and all results of in-class and out-of-class assignments and examinations are data sources for research and may be used in published research. All such use will always be anonymous.